Weekly Report
01/12/2015 - 01/18/2015

Jing Xia

January 18, 2015

1 Summary

This week I mainly focus on the data inspection project.

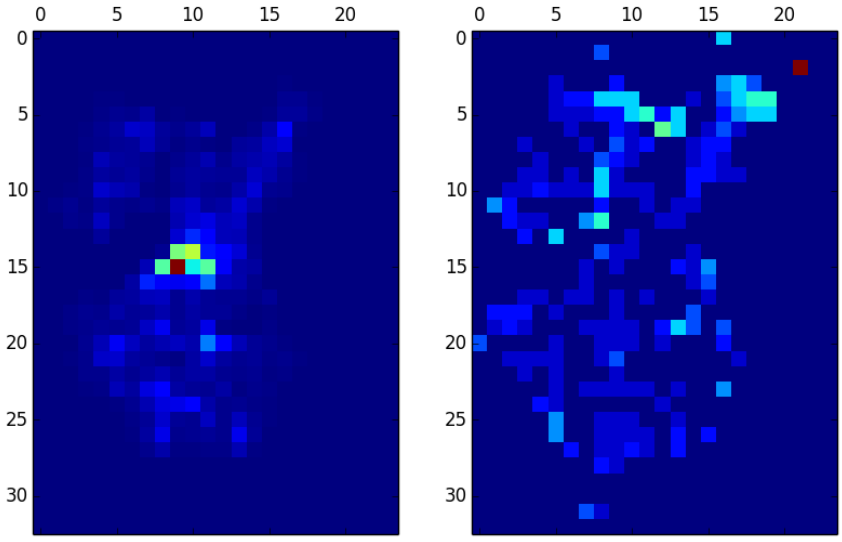
2 Projects

2.1 Project 1 - Rank Visualization

-

2.2 Project 2 - Data Inspection

This week I implemented **the mutual information** between two dimensions and visually compared the results with 2D-distribution of the underlying two dimensional datasets. Since it was just a test of the algorithm, I simply visualized it in Python. It turns out that highly correlated variables of two dimensions are not necessarily heavy distributed data regions, and vice versa. Referring Figure 1, for dimensions of latitude and longitude, data mostly distributes in the center (downtown area). However the point-wise mutual information matrix indicates higher correlated area at the upper right corner. It does make sense because, in this specific case, the higher correlated area has a distribution of $p(x = X_i, y = Y_i) = p(x = X_i) = p(y = Y_i)$. It infers such correlation that for (X_i, Y_i) is the only existing data for both row $x = X_i$ and column $y = Y_i$.
Also for **highly correlated local area discovery**, a simple method is adopted. The algorithm initially picks entries with top 5% values in the correlation matrix, expanding the areas to their 4-direction neighborhood if their neighboring entries are also highly correlated areas. The expansion continues until it meets a certain threshold (to be decided).
As mentioned in the dry-run talk, I need to take more flexible projection of views into consideration. And the mutual information computation needs to cover extra information such as user annotation.



(a) two dimensional distribution (b) point-wise mutual information

Figure 1: Comparison between data distribution and correlation distribution.

2.3 Project 3 - NBA Game Visualization

-

3 Paper Reading

-

4 Miscellaneous

1. Visual Data Inspection 2-page slides
2. VAST Challenge 2-page slides
3. Information visualization course project evaluation
4. Visual data inspection dry-run talk

5 To Do List

1. Data inspection project — projection.
2. NBA project discussion and data inspection project discussion.